UNIVERSITÉ DE MONTRÉAL

# A NEW ALGORITHM FOR SOLVING VARIATIONAL INEQUALITIES OVER POLYHEDRA, WITH APPLICATION TO THE TRAFFIC ASSIGNMENT PROBLEM

by

Patrice Marcotte

**CENTRE DE RECHERCHE SUR LES TRANSPORTS**

CASE POSTALE 6128, SUCCURSALE A, MONTRÉAL, QUÉBEC, H3C 3J7

# A NEW ALGORITHM FOR SOLVING VARIATIONAL INEQUALITIES OVER POLYHEDRA, WITH APPLICATION TO THE TRAFFIC ASSIGNMENT PROBLEM

by

Patrice Marcotte

Cette recherche a été rendue possible grâce à une subvention du C.R.S.H. (#410-81-0722-RL) , à une subvention du fonds F.C.A.C. (#83-AS-0026) ainsi qu'à une bourse de chercheur du Programme de soutien à l'emploi scientifique, Secrétariat à la science et à la technologie du Gouvernement du Québec.

Université de Montréal

Centre de recherche sur les transports - Publication #316
Ecole des Hautes Etudes Commerciales - Cahier du GERAD #8324

Septembre 1983

## ABSTRACT

The variational inequality problem can be formulated as a nonconvex, nondifferentiable optimization problem. We show that this formulation leads to a convergent algorithm that decreases the "gap" function in a monotone fashion.


- - - - - - - - - - - - - - -


## RÉSUMÉ

On peut formuler le problème d'inéquation variationnelle comme la minimisation d'une fonction non convexe et non différentiable. Nous montrons qu'il est possible de minimiser cette fonction à l'aide d'un algorithme de descente réalisable.


- - - - - - - - - - - - - - -


Key words : variational inequalities
nondifferentiable optimization
nonconvex programming
network optimization

## INTRODUCTION

Recently, much attention has been focused on the variational inequality formulation of the generalized traffic assignment problem with user-optimized behaviour : see Dafermos [5], Smith [20], Nguyen and Dupuis [17], Bertsekas and Gafni [3]. Although most solution methods proposed have only been tested in a transportation framework, the algorithms can readily be applied to a more general framework, including equilibrium programming, concave N- person games on polyhedra, provided that the cost function satisfies some monotonicity property. What distinguishes the multicommodity traffic assignment problem from the general variational inequality (VI) problem is the ease with which the linear approximations of the problem can be solved, namely by performing shortest path computations.

Despite its general applicability, our algorithm requires that the linear approximation subproblem be solved repeatedly and efficiently. The traffic assignment problem is a typical member of such a class of problems. In this paper, while presenting the results in a general setting, we will frequently stress the implications of our results when applied in a transportation context.

The paper will be organized as follows :

1. Formulation of the problem.
2. Properties of the "gap" function associated with a VI.
3. Presentation of two gap-minimizing algorithms based on nondifferentiable optimization techniques.

4. Convergence analysis.

5. Practical considerations regarding the implementation of the algorithm.

## 1. FORMULATION OF THE PROBLEM

We consider a continuous mapping $F : \Phi \to R^n$, where $\Phi$ is a nonempty convex compact subset of $R^n$. The VI problem associated with F and $\Phi$ consists in finding a vector x in $\Phi$, subsequently referred to as an equilibrium point, such that :

$$(x-y)^T F(x) \leq 0 \qquad \forall\, y \text{ in } \Phi \tag{1.1}$$

Existence of at least one solution is a direct consequence of Kakutani's fixed point theorem (see [8] ).

Definition 1.1

A mapping F is monotone (respectively strictly monotone) on $\Phi$ if : $(x-y)^T(F(x)-F(y)) \geq 0$ (respectively $> 0$) for all x,y in $\Phi$ $(x \neq y)$.

In the remainder of the paper, F is assumed to be monotone on $\Phi$.
If F is continuously differentiable, we have the following characterization :

Theorem 1.1. A continuously differentiable mapping F is monotone (respectively strictly monotone) on $\Phi$ if and only if its Jacobian matrix function $\nabla F$ satisfies :

$$(y-x)^T \nabla F(x)(y-x) \geq 0 \quad \text{(respectively } > 0) \quad \forall\, x,y \text{ in } \Phi\ (x \neq y).$$

Proof. See Auslender [2], theorem 1.2 of chapter VII. □

It is easily verified that monotonicity on $\Phi$ implies that the solution set of (1.1) is closed and convex. Moreover, monotonicity implies uniqueness of the solution.

The equilibrium traffic assignment problem can be formulated as (1.1), with $\Phi$ representing the set of multicommodity flow vectors satisfying the usual flow conservation equations and nonnegativity constraints, and F is a cost mapping representing congestion effects. Althoug $\Phi$ is not compact, flows can be restricted to a compact subset without modifying the solution set (see lemma 4.1.5 in Marcotte [14]).

Definition 1.2. The gap function associated with a VI problem is given by :

$$g(x) \triangleq \underset{y \in \Phi}{\text{Max}} \quad (x-y)^T F(x) \tag{1.2}$$

The gap function is always nonnegative. It is equal to zero if and only if x solves the VI problem. Therefore, solving (1.1) is equivalent to solving the following nondifferentiable optimization problem :

$$\underset{x \in \Phi}{\text{Min}} \; g(x) \tag{1.3}$$

In the next section we consider some important properties of g.

## 2. PROPERTIES OF THE GAP FUNCTION

The next results will be central to the algorithms of section 4.

Lemma 2.1. If F is Lipschitzian and $\Phi$ polyhedral, or if F is differentiable, then g if Lipschitzian.

Proof. If $\Phi$ is polyhedral, g can be written as the pointwise supremum of a finite number of Lipschitzian functions :

$$g(x) = \max_i (x-u^i)^T F(x)$$

where $u^i$ are the extreme points of $\Phi$ and is therefore Lipschitzian.

The differentiable case satisfies the hypothesis of proposition 4a in Rockafellar [18]. □

Lemma 2.2. The subdifferential $\partial g(x)$ is given by :

$$g(x) = \text{conv } \{F(x) + \nabla^T F(x)(x-y)\}$$

where $y \in \arg\min_{z \in \Phi} z^T F(x)$ and "conv" is the convex hull operator.

Proof. Straightforward application of the definition of generalized gradients (Clarke [4]) and the chain rule. □

Lemma 2.3. If F is continuously differentiable at x, then g is continuously differentiable at x if and only if the solution to the linear problem $\min_{y \in \Phi} y^T F(x)$ is unique.

Proof. See Hearn [11]. □

From the previous result we deduce that, if F is differentiable on $\Phi$ and $\Phi$ is strongly convex, then g is differentiable on $\Phi$.

The following lemma extends a corollary in Hearn [11] to the case where F is not differentiable.

Lemma 2.4. If the solution $\bar{y}$ to the linear problem $\underset{y\in\Phi}{\text{Min}}\ y^T F(x)$ is unique, then $d=\bar{y}-x$ is a feasible descent direction for g at x, provided that x is not an equilibrium point.

Proof. Let $z=x+\varepsilon(\bar{y}-x)$, where $\bar{y} \in \arg\min_{y\in\Phi} y^T F(x)$.

Consider the quantity : $q_\varepsilon = g(z) - g(x)$ representing the increase or decrease of the gap value in the direction $d=\bar{y}-x$. For small $\varepsilon$ we have :

$$\bar{y} \in \arg\min_{y\in\Phi} y^T F(x+\varepsilon(\bar{y}-x))$$

since F is continuous. Therefore :

$$\begin{aligned} q_\varepsilon &= (x+\varepsilon(\bar{y}-x)-\bar{y})^T F(x+\varepsilon(\bar{y}-x)) - (x-\bar{y})^T F(x) \\ &= (x-\bar{y})^T (F(x+\varepsilon(\bar{y}-x))-F(x)) + \varepsilon(\bar{y}-x)^T F(x+\varepsilon(\bar{y}-x)) \\ &= -\left(\frac{x+\varepsilon(\bar{y}-x)-x}{\varepsilon}\right)^T (F(x+\varepsilon(\bar{y}-x))-F(x)) \\ &\quad + \varepsilon(\bar{y}-x)^T F(x+\varepsilon(\bar{y}-x)) \end{aligned}$$

The first term is nonpositive by monotonicity of F, while the second term is negative for small enough $\varepsilon$, since g(x) is positive and F is continuous. This completes the proof. □

Definition 2.1. $\Gamma(x) \triangleq \arg\min_{y\in\Phi} y^T F(x)$.

Proposition 2.1. Let F be continuously differentiable and monotone on $\Phi$. If y is not an equilibrium point, then there exists a convex combination of extreme points of $\Gamma(x)$ :

$$c = \sum_{i=I}^{m} \lambda_i u^i \quad \text{such that :}$$

i) $d = c - x$ is a feasible descent direction for g at x.

ii) The directional derivative of g in the direction d satisfies :

$$\frac{d}{dt}\, g(x + t\,\frac{d}{\|d\|}) \leq -\frac{g}{\|d\|} \quad .$$

iii) $m \leq n+1$ .

Proof. 1. We first prove the proposition when $\Phi$, hence $\Gamma(x)$, is polyhedral. The gap function is written as :

$$g(x) = \max_{i=1,\ldots,\nu} (x-u^i)^T F(x) \tag{2.1}$$

where, as previously, $u^i$ are the extreme points of $\Phi$.

Following Danskin [6], the directional derivative of g in the direction d at x is given by the expression :

$$\frac{d}{dt}\, g(x+t\frac{d}{\|d\|}) = \max_{i\in I(x)} \frac{d^T}{\|d\|} (\nabla^T F(x)\cdot x + F(x) - \nabla^T F(x) u^i) \tag{2.2}$$

where : $i\in I(x) \Rightarrow g(x) = (x-u^i)^T F(x)$, i.e. $I(x)$ is the set of extremal points of $\Phi$ that are also extremal points of $\Gamma(x)$.

Let $A = \nabla F(x)$, giving :

$$Q \triangleq -\frac{d}{dt}\, g(x + \frac{t}{\|d\|}\, d)$$

$$= \frac{1}{\|d\|} \min_{i\in I(x)} \{ \sum_{j\in I(x)} \lambda_j (u^j-x)^T (A^T(u^i-x) - F(x)) \}$$

$$= \frac{1}{\|d\|} \min_{i\in I(x)} \{ \sum_{j\in I(x)} \lambda_j (u^j-x)^T A^T (u^i-x) \} + \frac{1}{\|d\|}\, g(x) \quad .$$

Since $(x-u^j)^T F(x) = g(x)$ $\forall$ $j \in I(x)$ and $\sum_{i \in I(x)} \lambda_i = 1$.

The proposition will be proved if there exists a nonnegative vector $\lambda$ such that :

$$\sum_{j \in I(x)} \lambda_j (u^j-x)^T A^T (u^i-x) \geq 0 \quad \forall \quad i \in I(x) \tag{2.3}$$

Define a matrix $B = (b_{ij})_{i,j \in I(x)}$ by :

$$b_{ij} \triangleq (u^j-x)^T A (u^i-x)$$

$$= (u^i-x)^T A^T (u^j-x).$$

(2.3) can be rewritten as :

$$\sum_{j \in I(x)} \lambda_j b_{ji} \geq 0 \qquad i \in I(x) \tag{2.4}$$

or, equivalently : $B^T \lambda \geq 0$. (2.5)

a. <u>The matrix B is positive semidefinite</u>.

Consider the symmetric matrix $C = B + B^T$. The elements of C are : $C_{ij} = (u^i-x)^T (A + A^T)(u^j-x)$. B will be positive semidefinite if and only il C is positive semidefinite.

Let $E = \{u-x | u \in \Phi\}$ and V be the vector space spanned by elements of E. We will show that $A+A^T$ is positive semidefinite on V. By theorem 1.1, one has :

$$(y-x)^T (A+A^T)(y-x) \geq 0 \quad \forall \quad y \in \Phi \quad . \tag{2.6}$$

Let $y = \alpha(u^1-x) + \beta(u^2-x)$ with $u^1$ and $u^2$ in $\Phi$.

Then : $y = (\alpha+\beta)\left[\left(\frac{\alpha}{\alpha+\beta}u^1 + \frac{\alpha}{\alpha+\beta}u^2\right) - x\right]$

$= (\alpha+\beta)(u-x)$ where $u \in \Phi$ by convexity of $\Phi$.

Now : $y^T(A+A^T)y = (\alpha+\beta)^2(u-x)^T(A+A^T)(u-x) \geq 0$ by (2.6).

We define the scalar product : $\langle v^1, v^2 \rangle = v^{1T}(A+A^T)v^2$ on the Hilbert space V.*

For any z in $R^{|I(x)|}$ one has :

$$z^TCz = \sum_{i,j \in I(x)} \langle v^i, v^j \rangle z_i z_j$$

$$= \sum_{i,j \in I(x)} \langle z_i v^i, z_j v^j \rangle$$

$$= \left\| \sum_{i \in I(x)} z_i v^i \right\|^2_{A+A^T} \geq 0$$

Hence C is positive semidefinite

=> B is positive semidefinite.

b. Feasibility of the system $\{B^T\lambda \geq 0 \quad \lambda \geq 0\}$

Consider the following pair of primal and dual linear programming problems :

---

* Actually, C is the Gram matrix associated with the scalar product $\langle .,. \rangle$ .

<u>Primal</u>

$$\min_{\lambda} \; 0^T\lambda$$

$$\text{s.t. } B^T\lambda \geq 0 \qquad (2.7)$$

$$\lambda \geq 0$$

$$\sum_{i\in I(x)} \lambda_i = 1$$

<u>Dual</u>

$$\max_{z,\alpha} \; \alpha$$

$$\text{s.t. } Bz + \alpha \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \leq 0 \qquad (2.8)$$

$$z \geq 0$$

The dual problem can be rewritten as

$$\max_{z,\alpha} \; \alpha$$

$$\text{s.t.} \quad Bz \leq -\alpha \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \qquad (2.9)$$

$$z \geq 0$$

Suppose $\alpha$ is positive. This implies that z is non zero and :

$$z^T Bz \leq -\alpha \sum_{i\in I(x)} z_i < 0 \; .$$

This contradicts the positive semidefinitesess of B. Hence the dual admits a finite solution. By the linear programming duality theorem, the primal must be feasible. This completes the proof of points 1 and 2. Point 3 results from a straightforward application of Carathéodory's theorem (see Rockafellar [18]) which states that any point of a convex bounded polyhedron in $R^n$ can be expressed as convex combination of at most n+1 of its extreme points.

2. <u>Nonpolyhedral case</u>

Let $\{u^i\}_{i=1,\ldots,u_n}$ be a "dense" set of extreme points of $\Gamma(x)$ such

that, for any other extremal point u in $\Gamma(x)$ one has : $\|u-u^i\| \leq \frac{1}{n}$ for at least one index i. For each finite set $\{u^i\}_{i=1,\ldots,\nu_n}$ we associate a nonnegative weighting vector $\lambda$ such that the constraints of (2.7) are satisfied. Repeating the previous derivation we get :

$$Q \triangleq \text{directional derivative in direction d at x}$$

$$= \frac{1}{\|d\|} \min_{\substack{u\in\Gamma(x) \\ u \text{ extreme} \\ \text{point of } \Phi}} \{ \sum_{j=1}^{\nu_n} \lambda_j (u^j-x)^T A^T (u-x) \} + \frac{1}{\|d\|} g(x) \qquad (2.10)$$

For each extreme point of $\Gamma(x)$ one has :

$$R \triangleq \sum_{j=1}^{\nu_n} \lambda_j (u^j-x)^T A^T (u-x)$$

$$= \sum_{j=1}^{\nu_n} \lambda_j (u^j-x)^T A^T (u^i-x) + \sum_{j=1}^{\nu_n} (u^j-x)^T A^T (u-u^i)$$

$$\text{with } \|u-u^i\| \leq \frac{1}{n} .$$

Therefore : $R \geq \frac{1}{n} \|u^j-x\| \cdot \|A^T\| \geq -\frac{\beta}{n}$

where $\beta = \sup_{z^1,z^2,z^3 \in \Phi} \|z^1-z^2\| \cdot \|\nabla^T F(z^3)\| < \infty$ .

Let $c_n = \sum_{i=1}^{u_n} \lambda_i u^i$, $d_n = c_n - x$, and $\bar{c} \in \Gamma(x)$ be any cluster point of the sequence $\{c_n\}_n$ and $\bar{d} = \bar{c}-x$. We have that the directional derivative $d(c_n)$ of g in direction $d_n$ at x is less than $-\frac{g}{\|d_n\|} + \frac{\beta}{n}$ . By a continuity argument we obtain : $d(c) = \lim_{n\to\infty} d(c_n) \leq -\frac{g}{\|d\|}$ .

Point 3 is again proved by invoking Carathéodory's theorem. □

The proof and results contained in proposition 1 call for some comments :

i) If $\Gamma(x)$ is polyhedral (this is the case if $\Phi$ is polyhedral) and its extreme points were known a priori, it would be possible to find a steepest descent direction by solving the linear programming problem :

$$\begin{aligned} &\underset{\lambda,\gamma}{\text{Max}}\ \gamma \\ \text{s.t.}\quad &B\lambda \geq \gamma \\ &\sum_{i\in I(x)} \lambda_i = 1 \end{aligned} \tag{2.11}$$

ii) If only a subset of extreme points are known, and the directional derivative is more than $-\frac{q}{\|d\|}$, it is possible to improve the direction by incorporating into it the most "troublesome" element of $\Gamma(x)$, i.e. an extreme point $u^k$ of $\Gamma(x)$ such that :

$$u^k \in \underset{i \in I(x)}{\arg\min} \sum_{i\in I(x)} \lambda_i (u^i - x)^T A (u^k - x) \tag{2.12}$$

Let $c^T = \sum_{i\in I(x)} \lambda_i (u^i - x^k)^T A$; $u^k$ is then solution of the linear problem :

$$\min_{y\in\Gamma(x)} c^T y \tag{2.13}$$

and a new direction can be computed by including the index k into the set $\bar{I}(x) \subseteq I(x)$ of known extreme points of $\Gamma(x)$, and solving (2.11) with $\bar{I}(x)$ in place of $I(x)$.

iii) The method used for finding a gap reducing direction resembles methods of nondifferentiable optimization for finding descent directions, such as $\varepsilon$-conjugate subgradients, or projection of the origin on a set of previously generated subgradients. In our case, "conjugacy" is defined with respect to the scalar product induced by the symmetrized Jacobian matrix $\nabla F(x) + \nabla^T F(x)$.

iv) Since g admits computable descent directions outside its solution set, it is unimodal and has no stationary points(*). One might imagine that in such circomstances, the gap function is at least quasiconvex. We show by a counterexample that g may be non quasiconvex.

Example 2.1

$$\text{Let } \Phi = \{x \in R^2 \mid \begin{matrix} x_1, x_2 \geq 0 \\ x_1 + x_2 \leq 1 \end{matrix}\}$$

$$\text{and } F(x_1, x_2) = \begin{cases} x_1^2 + 4x_1x_2 \\ x_2^2 + 4x_1x_2 \end{cases} \qquad x^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

We have, on $\Phi$ : $g(x) = (x-0)^T F(x) = x^T F(x)$

$$\nabla F = \begin{bmatrix} 2x_1 + 4x_2 & 4x_1 \\ 4x_2 & 2x_2 + 4x_1 \end{bmatrix}$$

which is positive semidefinite for every x in $\Phi$. Therefore G is monotone.

However :

$$g\begin{pmatrix} 1 \\ 0 \end{pmatrix} = g\begin{pmatrix} 0 \\ 1 \end{pmatrix} = 1 \quad \text{but} \quad g\begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} = \frac{5}{4} > 1.$$

---

(*) Our definition of a stationary point for nonconvex nondifferentiable functions follows Rockafellar [19] : it is a point where no feasible descent direction exist. Mifflin [16] gives a different definition.

## 3. A GAP-MINIMIZING ALGORITHM

In this section we present a new descent algorithm that minimizes directly the gap function, using nonstandard nondifferentiable optimization methods. The algorithm is based on the results of proposition 2.1. However, since the explicit knowledge of all extreme points of $\Gamma(x)$ would be prohibitive, we generate new extreme points of $\Gamma(x)$ only when required. Two versions of the basic algorithm will be given. Convergence analysis and implementation will be discussed in sections 4 and 5 respectively.

In the remainder of the paper we assume that $\Phi$ is polyhedral, and F is twice continuously differentiable. The nonpolyhedral case will be considered in a forthcoming paper.

We denote by LP(x) any extremal solution of the problem :

$$\min_{y \in \Phi} y^T F(x) \ .$$

Algorithm 1 (Exact line search)

1. Set k to zero. Choose $\varepsilon > 0$. ($\varepsilon$ small).

   Let $x^{(1)}$ be any feasible point of $\Phi$.

2. $k \leftarrow k+1$

   $u^1 \leftarrow LP(x^k)$

   $g^k \leftarrow g(x^k) = (x^k - u^1)^T F(x^k)$

   $\lambda_1 \leftarrow 1$

   $m \leftarrow 1$ .

3. $d \leftarrow u^1 - x^k$

4. Let $u^{m+1}$ be an extremal solution of (2.13) :

$$\min_{y \in \Gamma(x)} c^T y$$

If there exists i ($0 \leq i \leq m$) such that $u^{m+1} = u^i$ go to 6.

5. Find a steepest descent direction

$$d = \sum_{i=1}^{m+1} \lambda_i u^i - x$$

by solving (2.11).

$m \leftarrow m+1$ and return to 3.

6. (Line search) $\bar{\theta} \leftarrow \arg \min_{0 \leq \theta \leq 1} g(x^k + \theta d)$

$$x^{k+1} \leftarrow x^k + \bar{\theta} d \ .$$

7. (Termination test)

If $g(x^{k+1}) < \varepsilon$, STOP otherwise return to 2. $\Delta$

In the case of the traffic assignment problem, step 4 of algorithm 1 can be efficiently solved by a modified shortest path routine that retains only those shortest paths that are optimal (among shortest paths) with respect to the c-criterion.

When a point $u^i$ is repeated (step 4), this means that $\min_{y \in \Gamma(x)} c^T y = 0$ and that consequently the descent direction has slope better or equal to $-\frac{g}{|d|}$. The criterion could be weakened somewhat by only requiring that :

$$c^T y \geq -\frac{g}{\mu} \text{ for some number } \mu > 1 \ .$$

Although algorithm 1 is not practical because of the requirement for exact line searches, it illustrates the main ideas that lead to the next algorithm. Algorithm 2 is motivated by the concepts of $\varepsilon$-subdifferential [13] and of sufficient decrease of the objective function. Whenever minimization along a direction does not provide a sufficient gap decrease, we look for a better direction involving the gradient of g at a nearby point $x+\varepsilon$, which "always" exists since g is differentiable almost everywhere.

Before giving a formal description of algorithm 2, let us consider the application of algorithm 1 to a theoretical example designed by Wolfe [22] to demonstrate the poor convergence properties of the Frank-Wolfe algorithm in a neighborhood of the solution.

Example 3.1

$$F(x) = \begin{cases} x_1 \\ x_2 \end{cases}$$

$$\Phi = \left\{ x \in R^2 \,\middle|\, \begin{array}{l} x_2 - 2x_1 \leq 2 \\ 2x_1 + x_2 \leq 2 \end{array} \quad \text{and} \quad x_1, x_2 \geq 0 \right\}$$

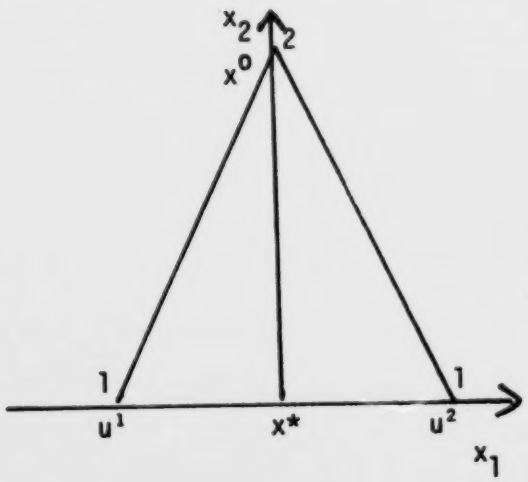


Figure 3.1

We take $x^0 = \begin{pmatrix}0\\2\end{pmatrix}$ as starting point; $x^* = \begin{pmatrix}0\\0\end{pmatrix}$ is the equilibrium point.

For $x_1 \geq 0$, the gap is given by :

$$g(x) = (x-u^1)^T F(x)$$

where $u^1 = \begin{pmatrix}-1\\0\end{pmatrix}$. For $x^1 \leq 0$ it is given by :

$$g(x) = (-u^2)^T F(x)$$

where $u^2 = \begin{pmatrix}1\\0\end{pmatrix}$.

At $x^0$ we decide, arbitrarily, to take $u^2$ as solution of the problem :

$$\underset{y\in\Gamma}{\text{Min}}\ y^T F(x^0) \quad .$$

We have : $g(x^0) = \left[\begin{pmatrix}0\\2\end{pmatrix} - \begin{pmatrix}-1\\0\end{pmatrix}\right]^T \begin{pmatrix}0\\2\end{pmatrix} = 4$ .

For a given x, consider the gap-minimizing $\bar{x}$ obtained by moving in the direction of a shortest path. If $x_1 \geq 0$ we have :

$$\bar{x} = \left[\begin{pmatrix}(1-\lambda)x_1-\lambda\\(1-\lambda)x_2\end{pmatrix} - \begin{pmatrix}1\\0\end{pmatrix}\right]^T \cdot \begin{bmatrix}(1-\lambda)x_1-\lambda\\(1-\lambda)x_2\end{bmatrix}$$

$$= [\lambda(1+x_1)-x_1]^2 + (\lambda x_2 - x_2)^2 + \lambda(x_1+1) - x_1$$

$$= \lambda^2 [(1+x_1)^2 + (x_2)^2] + \lambda[-2x_1(1+x_1) - 2(x_2)^2 + x_1 + 1]$$

which is minimized by :

$$\lambda = \frac{2(x_1)^2+2(x_2)^2+x_1-1}{2(x_1)^2+2(x_2)^2+4x_1+2} \qquad (3.1)$$

$\lambda < 0$ means that we cannot cross the vertical axis; in this case we set :

$\lambda = \frac{x_1}{1+x_1}$ . Thus :

$$\lambda = \text{Max}\left\{\frac{x_1}{1+x_1}\ ,\ \frac{2(x_1)^2+2(x_2)^2+x_1-1}{2(x_1)^2+2(x_2)^2+4x_1+2}\right\}. \tag{3.2}$$

For $x_1 \leq 0$, we get, by a symmetric argument :

$$\lambda = \text{Max}\left\{\frac{|x_1|}{1+|x_1|}\ ,\ \frac{2(x_1)^2+2(x_2)^2+|x_1|-1}{2(x_1)^2+2(x_2)^2+4|x_1|+2}\right\}\ . \tag{3.3}$$

From (3.3) we derive :

$$x^0 = \begin{pmatrix}0\\2\end{pmatrix} \qquad g(x^0) = 4$$

$$\lambda_1 = \text{Max}\{\tfrac{7}{10},0\} = \tfrac{7}{10} \qquad x^1 = \begin{pmatrix}-7/10\\6/10\end{pmatrix} \quad g(x^1) = \tfrac{155}{100}$$

$$\lambda_2 = \text{Max}\{\tfrac{14}{55},\tfrac{7}{17}\} = \tfrac{7}{17} \qquad x^2 = \begin{pmatrix}0\\6/17\end{pmatrix} \quad g(x^2) = \tfrac{36}{289}$$

Now, at $x^2$, there are two LP-solutions; any descent direction for g at $x^2$ must take both into account. The steepest descent direction obtained by solving (3.11) is :

$$d = x^2 - x^*$$

and minimization of the gap along d yields the solution.

It is instructive to compare the first step of the algorithm with the Frank-Wolfe step for solving the equivalent optimization problem :

$$\text{Min}_{x\in\Gamma}\ \frac{1}{2}\|x\|^2 \tag{3.4}$$

The first Frank-Wolfe iteration would have given :

$$\bar{\lambda}_1 = \frac{4}{5} \quad \text{and} \quad \bar{x}^1 = \begin{pmatrix} -4/5 \\ 2/5 \end{pmatrix}$$

whit $g(\bar{x}^1) = \frac{8}{5} > \frac{155}{100} = g(x^1)$.

It will be true in general that, when they are comparable, the stepsize of algorithm 1 will be smaller than the Frank-Wolfe stepsize.

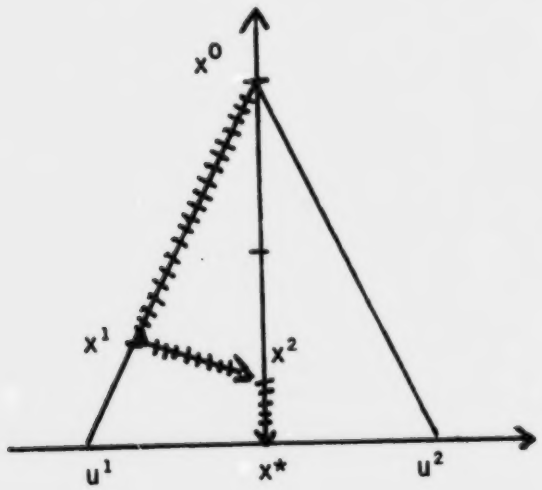


Figure 3.2 - Sequence generated by algorithm 1

We now present the second version of the basic algorithm.

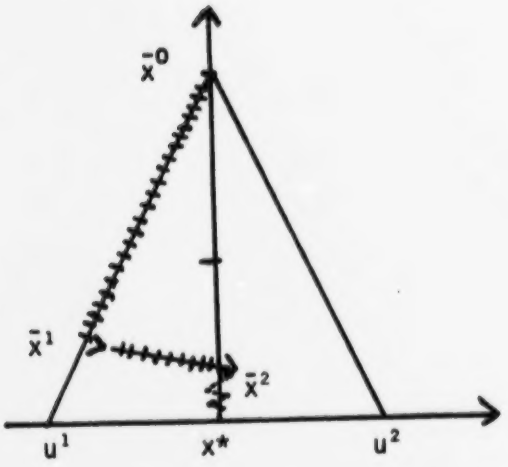


Figure 3.3 - Sequence generated by Frank-Wolfe algorithm

Algorithm 2

1. $k \leftarrow 0$. Let $x^{(1)}$ be any feasible point in $\Phi$ .

2. $k \leftarrow k+1$ $\quad m \leftarrow 1$

   $u^1 \leftarrow LP(x^k)$

   $g_k \leftarrow g(x^k) = (x^k - u^1)^T F(x^k)$

   If $g_{k+1} < \varepsilon$ (precision parameter) STOP.

3. $d \leftarrow u^1 - x^k$

4. $\bar{x} \leftarrow x^k + \alpha g_k d$ where $\alpha$ is a predetermined positive number (see section 5)

5. If $g(\bar{x}) \leq g_k (1 - \alpha g_k)$ then

   perform a line search in direction d to find $x^{k+1}$; set $g_{k+1} \leftarrow g(x^{k+1})$

   and return to step 2.

   Otherwise :

6. $u^{m+1} \leftarrow LP(\bar{x})$

   $m \leftarrow m + 1$

7. Find a new direction $d = \sum_{i=1}^{m} \lambda_i u^i - x^k$ where is solution of the linear program :

$$\max_{\lambda,\gamma} \gamma$$

$$\text{s.t. } B^T\lambda \geq \gamma \tag{3.5}$$

$$\sum_{i=1}^{m} \lambda_i = 1 \ , \ \lambda \geq 0 \quad \text{and return to step 4.} \qquad \Delta$$

The existence proof for the parameter $\alpha$ at step 4 is referred to section 5, side along with the convergence proof. The nonnegativity of the optimum in (3.5) is a direct consequence of proposition 1. At step 5, any valid line search can be used (see Armijo [1], Huard [12], Wolfe [21]).

When F is a gradient mapping, algorithm 2 closely resembles the Frank-Wolfe method and the restriction variant of Wolfe as implemented by Fratta and Gerla [9] or Florian [7] . In this variant, the descent direction takes into account a fixed number of extreme points previously generated. It is conceivable that this method is more apt to reduce the gap that the standard Frank-Wolfe step.

Remark. Our method is not equivalent to a nondifferentiable Frank-Wolfe method applied to g. When g is differentiable at x, the Frank-Wolfe direction is given by $d = y - x$, where :

$$y \in \arg\min_{z\in\Phi} z^T [F(x) + \nabla^T F(x)(x-u)] \tag{3.6}$$

with $u = \arg\min_{s\in\Phi} s^T F(x)$. The vector u is not necessarily solution of (3.6).

Because g is locally Lipschitzian, it is possible to apply an algorithm of Mifflin [15] to reduce the gap. However his approach requires projections on $\Phi$ and would therefore be quite inefficient.

## 4. CONVERGENCE ANALYSIS

Convergence of algorithm 2 will be achieved if :

i) A positive number $\alpha$ can be found such that the criterion at step 5 is satisfied after a finite number of trials, and one eventually returns to step 2.

ii) A sequence satisfying (15) converges to zero.

The first point is a result of the following two lemmas.

Lemma 4.1. There exists a positive number $\alpha_1$ smaller than one such that, at step 4 of algorithm 2, one has :

$$\alpha \le \alpha_1 \quad \text{and} \quad y = LP(\tilde{x}) \Rightarrow (x^k - y)^T F(x^k) \ge \frac{1}{2} \cdot g_k .$$

Proof : Let $\alpha = \sum_{i=1}^{m} \lambda_i u^i - x^k$ as before, $\tilde{x} = x^k + \alpha_1 g_k d$ and $\varepsilon = \alpha_1 g_k$.
Using the Taylor expansion of F at $x^k$ we derive :

$$(y-x^k)^T F(x^k) = (y-\tilde{x}+\varepsilon d)^T F(x-\varepsilon d)$$

$$= (y-\tilde{x}+\varepsilon d)^T (F(\tilde{x}) - \varepsilon A^T d + \frac{1}{2}\varepsilon^2 d^T B d)$$

where $A = \nabla^T F(\tilde{x})$, and B $\nabla^2 F(\xi)$ is the three dimensional array of second

partial derivatives of F evaluated at a point $\xi$ on the segment $[x^1, \tilde{x}]$.

We can assume that $g(\tilde{x}) = (y-\tilde{x})^T F(\tilde{x}) \leq -\frac{2}{3} g_k$; otherwise the gap decrease would have been significant. Thus :

$$(y-x^k)^T F(x^k) \leq -\frac{2}{3} g_k + K\varepsilon$$

where $K = \text{Sup} \left| (y-\tilde{x})^T(-A^T d + \frac{1}{2} \varepsilon d^T B d) + d^T(F(\tilde{x}) - \varepsilon A^T d + \frac{1}{2} \varepsilon^2 d^T B d) \right| < \infty$

since all functions involved are continuous and $\Phi$ is compatt.

Taking $\alpha_1 = \frac{1}{6K}$ gives :

$$(y-x^k)^T F(x^k) \leq -\frac{2}{3} g_k + \frac{1}{6K} K g_k$$

$$= -\frac{1}{2} g_k \, . \qquad \square$$

Lemma 4.2. There exists a positive number $\alpha_3 \leq \alpha_1$ such that, for all $\alpha$ less or equal to $\alpha_3$, the criterion of step 5 of algorithm 2 must be met after a finite number of trials.

Proof. Since the number of extreme points is finite, either the criterion is satisfied, either some extreme point is repeated. Let $u^p (1 \leq p < n)$ be such a point. Consider any point x lying on the segment $[x^k, \bar{x}=x^k+\alpha_2 g_k d]$ where $\alpha_2 \leq \alpha_1$ (Recall that $d = \sum_{i=1}^{m} \lambda_i u^i - x^k$). There exists $\theta (0 \leq \theta \leq 1)$ such that $x = x^k + \theta d$.

We have : $u^1 = LP(x^k)$, $u^p = LP(\bar{x})$ and we set $u = LP(x)$; we write :

$$0 \geq (u^p - u)^T F(\bar{x})$$

$$\geq (u^p - u)^T (F(x) + \theta \alpha_2 g_k A d) - \tilde{K}(\theta \alpha_2 g_k)^2$$

where $A = \nabla^T(F(x^k))$ and $\tilde{K}$ depends on the modulus of continuity of $F, \nabla^T F$ and $\nabla^2 F$.

Since $(u^P-u)^T F(x) \geq 0$ by definition of $u^P$, we obtain :

$$\tilde{K}(\theta\alpha_2 g_k)^2 \geq ((u^P-x^k) + (x^k-u))^T \theta\alpha_2 g_k Ad$$

$$\geq \theta\alpha_2 g_k (x^k-u)^T Ad$$

by construction of the direction d (step 7 of algorithm 2).

Thus : $(u-x^k)^T Ad \geq -(\tilde{K}\theta\alpha_2) g_k$ and :

$$(u-x)^T Ad \geq -(\tilde{K}\theta\alpha_2) g_k + (x^k-x)^T Ad$$

$$= (-\tilde{K}\theta\alpha_2 - \theta\alpha_2 d^T Ad) g_k$$

$$\geq -g_k/3 \quad \text{for a suitably small } \alpha_2.$$

The directional derivative of g along the segment considered is given by the expression :

$$\frac{d}{dt}\, g\left(x + t\frac{d}{\|d\|}\right) = \frac{1}{\|d\|}\left[\max_{y=LP(x)} (x-y)^T Ad - g(x)\right]$$

$$\leq \frac{1}{\|d\|}\left[\max_{y=LP(x)} (x-y)^T Ad - \frac{1}{2}\, g_k\right]$$

by lemma 4.1

$$\leq \frac{1}{\|d\|}\left(\frac{1}{3}\, g_k - \frac{1}{2}\, g_k\right) = -\frac{g_k}{6\|d\|}$$

$$\leq -\frac{1}{6S}\, g_k \quad \text{where} \quad S = \sup_{u,x\in\Phi} \|u-x\| < \infty .$$

It follows that :

$$g(\bar{x}) - g_k \leq -\alpha_2 g_k \frac{1}{6S} g_k$$

$$\Rightarrow \qquad g(\bar{x}) \leq g_k\left(1 - \frac{\alpha_2}{6S} g_k\right) .$$

To complete the proof, take $\alpha_3 = \frac{\alpha_2}{6S}$ . □

Remark

Lemmas 4.1 and 4.2 could have been proven with the slightly weaker requirement that $\nabla F$ be Lipschitzian rather than twice continuously differentiable.

Proposition 4.1

Any cluster point of the sequence $\{x^n\}_n$ generated by algorithm 2 is an equilibrium point.

Proof. Consider the real sequence $\{\zeta_n\}$ defined by the recursion :

$$\zeta_{n+1} = (1-a\zeta_n)\zeta_n \quad \text{where} \quad a \leq \frac{1}{\zeta_1} \quad . \tag{4.2}$$

Since this sequence is monotonically decreasing, it must have, for any $\zeta_1 \leq \frac{1}{a}$ , a limit point. Let $\zeta^*$ be this point. Then, by taking limits, we obtain :

$$\zeta^* = (1-a\zeta^*)\zeta^*$$

$$\Rightarrow \zeta^* = 0 .$$

Now the sequence $g(x^n)$ is bounded above by a sequence of the form (4.2). Thus : $\lim_{n\to\infty} g(x^n) = 0$. Finally continuity of g yields the derired result. □

## 5. IMPLEMENTATION CONSIDERATIONS

In this section we discuss a possible efficient implementation of algorithm 2 for solving large-scale traffic assignment problems. This class of problems is characterized empirically by the small number of extremal points (set of paths) included in the equilibrium solution. Therefore the number m of algorithm 2 is likely to be small.

Also, in most instances, any extreme point will give a descend direction of good quality. Far from the equilibrium, it is advisable to bypass completely step 4, and to replace step 5 by :

$$x^{k+1} \leftarrow x^k + \alpha_k d$$

where $\alpha_k$ is a "pseudo" Frank-Wolfe stepsize satisfying :

$$(x^{k+1}-u^1)^T F(x^k-\alpha(u^1-x^k)) = 0 \ .$$

Although $x^{k+1}$ is not gap-minimizing, it heuristically produces good starting iterations. After some iterations, it will happen that $g(x^{k+1}) > g(x^k)$; from that point on we can, always keeping the "pseudo" Frank-Wolfe direction $x-u^1$, perform a regular line search, demanding a sufficient decrease of the gap. After a second group of iterations this will no more be possible. We will then switch to the original procedure. At this point the gap will most probably be already very small.

It is possible to improve procedure at a very low extra cost by performing a "pseudo" PARTAN step along the direction $x^{k+1}-x^k$. This has proved very successful for the separable case (see Guélat and Florian [10]).

Applied to the example of section 3, this procedure would yield the sequence shown in Figure 5.1.

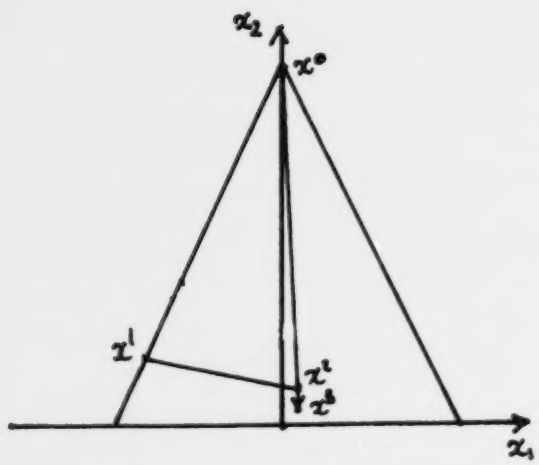


Figure 5.1

---

REMERCIEMENTS

J'aimerais remercier les professeurs Richard Loulou
et Michel Minoux pour des discussions fructueuses.

- - - - - - - - - - - - - - -

## REFERENCES

[1] ARMIJO, L. (1966), Minimization of functions having continuous partial derivatives. Pacific Journal of Mathematics 16.

[2] AUSLENDER, A. (1976), Optimisation; méthodes numériques. Masson, Paris.

[3] BERTSEKAS, D., GAFNI, E.M. (1982), Projection methods for variational inequalities with application to the traffic assignment problem. Math. Prog. Study 17.

[4] CLARKE, F.H. (1975), Generalized gradients and applications. Trans. Amer. Math. Soc. 205.

[5] DAFERMOS, S.C. (1980), Traffic equilibrium and variational inequalities. Transportation Science 14.

[6] DANSKIN, J.M. (1966) The theory of max-min, with applications. SIAM Journal of Applied Mathematics 14, 4.

[7] FLORIAN, M. (1977), An Improved Linear Approximation Algorithm for the Network Equilibrium (Packet Switching) Problem. Proceedings of the IEEE Conference on Decision and Control.

[8] FRIEDMAN, J.W. (1977), Oligopoly and the Theory of Games. North-Holland, Amsterdam.

[9] GERLA, M., CANTOR, D.G. (1974), Optimal routing in a packet-switched computer network. IEEE Trans on Computers, C-23, Vol. 10.

[10] GUÉLAT, J. (1983), Algorithmes pour le problème d'affectation du trafic d'équilibre avec demandes fixes. Comparaison. Publication #299, Centre de recherche sur les transports, Université de Montréal.

[11] HEARN, D.W. (1981), The gap function of a convex program. Operational Research Letters 1, 2.

[12] HUARD, P. (1981), Une approche synthétique des recherches linéaires dans les méthodes de pente. E.D.F. Bulletin de la direction des études et recherches, Paris, France.

[13] LEMARÉCHAL, C. (1978), Nonsmooth Optimization and Descent Methods. RR-78-4. IIASA, Laxenburg, Austria.

[14] MARCOTTE, P. (1981). Design optimal d'un réseau de transport en présence d'efforts de congestion. Ph.D. thesis, dép. d'informatique et de recherche opérationnelle, Université de Montréal, Montréal, Canada.

[15] MIFFLIN, R. (1982), A modification and an extension of Lemaréchal's algorithm for nonsmooth minimization. Math. Prog. Study 17.

[16] MIFFLIN, R. (1977), An algorithm for constrained optimization with semismooth functions. Math. of Op. Res. 2, 2.

[17] NGUYEN, S., DUPUIS, C. (1981), An Efficient Method for Computing Traffic Equilibria in Networks with Asymmetric Transportation Costs. Publication #205, Centre de recherche sur les transports, Université de Montréal.

[18] ROCKAFELLAR, R.T. (1970), Convex Analysis. Princeton University Press, Princeton, N.J.

[19] ROCKAFELLAR, R.T. (1981), La théorie des sous-gradients et ses applications à l'optimisation. Presses de l'Université de Montréal, Montréal, Canada.

[20] SMITH, M.J. (1979), The existence, uniqueness and stability of traffic equilibrium. Transportation Research B 13B, 4.

[21] WOLFE, P. (1962), Convergence conditions for ascent methods. SIAM Review 11, 2.

[22] WOLFE, P.(1970), Convergence In Nonlinear Programming. In : J. Abadie, ed.: Integer and Nonlinear Programming, North-Holland, Amsterdam.